

Syddansk Universitet

## Can semiquantitative measurements of SUVmax and cut-off values differentiate colorectal malignant from benign lessions?

Nguyen, Tram; Hess, Søren; Petersen, Henrik ; Alavi, Abass; Høilund-Carlsen, Poul Flemming

*Published in:*  
Hellenic Journal of Nuclear Medicine

*DOI:*  
[10.1967/s002449910551](https://doi.org/10.1967/s002449910551)

*Publication date:*  
2017

*Document version*  
Publisher's PDF, also known as Version of record

*Document license*  
Unspecified

*Citation for pulished version (APA):*  
Nguyen, T., Hess, S., Petersen, H., Alavi, A., & Høilund-Carlsen, P. F. (2017). Can semiquantitative measurements of SUVmax and cut-off values differentiate colorectal malignant from benign lessions? Hellenic Journal of Nuclear Medicine, 20(2), 113-121. DOI: 10.1967/s002449910551

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Can semiquantitative measurements of SUVmax and cut-off values differentiate colorectal malignant from benign lesions?

Tram Nguyen<sup>1</sup> MSc, PhD,  
Søren Hess<sup>1,2</sup> MD,  
Henrik Petersen<sup>1</sup> MD,  
Abass Alavi<sup>3</sup> MD, MD(Hon.),  
PhD(Hon.), DSc(Hon.),  
Poul Flemming Høilund Carlsen<sup>1</sup>  
MD, DMSci

1. Department of Nuclear Medicine,  
Odense University Hospital,  
Odense, Denmark.

2. Department of Radiology and  
Nuclear Medicine, Hospital of  
Southwest Jutland, Esbjerg,  
Denmark.

3. Division of Nuclear Medicine,  
Hospital of the University of Pennsylv-  
ania, Philadelphia, Pennsylvania,  
USA.

**Keywords:** Quantitative analysis,  
-Positron emission tomography,  
-Computer assisted diagnosis,  
-Prognosis, -Colorectal cancer

## Corresponding author:

Tram Nguyen MSc, PhD,  
Department of Nuclear Medicine,  
Odense University Hospital,  
Sdr. Boulevard 29, DK-5000  
Odense, Denmark.  
tram.quang.nguyen@rsyd.dk

## Received:

30 May 2017

## Accepted:

8 June 2017

## Abstract

**Objectives:** We investigated maximum standardized uptake value (SUVmax) and cut-off values for differentiation of malignant and benign lesions in colorectal cancer (CC) as multiple studies have questioned their validity. We also investigated more extended indices using common semi-quantification analysis in incidental colorectal findings (ICF). **Subjects and Methods:** Fluorine-18-fluoro deoxy glucose positron emission tomography/computed tomography in 25 patients with a total of 30 focal ICF was retrospectively analysed using dedicated software. Method variability was tested through application of three common threshold-based lesion delineation techniques as well as a partial-volume correction (PVC). Lesion SUVmax, SUVmean, metabolically active volume (MAV) and mean total lesion glycolysis (TLG) were thereby extracted along with PVC corrected values (cSUVmean, cTLG) and SUVpeak. **Results:** In all lesions, SUVmax was  $>5$  and  $\text{SUVmean} \geq 2.7$ . Malignant SUVmax values ( $\text{mean} \pm \text{SD}$ :  $16.5 \pm 6.2$ ) were overall significantly higher than benign levels ( $9.8 \pm 3.6$ ). There was a substantial overlap with values in polyps/adenomas ( $14.4 \pm 7.7$ ). Both SUVpeak and SUVmean showed similar characteristics. Malignant MAV and TLG showed more distinct levels. Though different segmentation methods introduced variations, largest in MAV ( $-58.6\%$ - $141.5\%$ ), and PVC generally increased measures significantly by a factor of 1.2-2.7, neither changed relative levels much. SUVmax values were inadequate for aetiological differentiation of ICF, which also precludes a clinically significant cut-off value. The same applies to SUVpeak and SUVmean while TLG measures may be more indicative. **Conclusion:** Semi-quantitative measurements of SUVmax and cut-off values proved inadequate for differentiating colorectal malignancies from benign findings. While integrated measures, e.g. cTLG, are potentially better indicators of disease severity and extent, more optimal segmentation and PVC methods are required.

*Hell J Nucl Med* 2017; 20(2): 113-121

*Epub ahead of print:* 12 July 2017

*Published online:* 8 August 2017

## Introduction

Semi-quantitative measures for soft tissue mass classification have been explored in oncologic positron emission tomography (PET) studies using 2-deoxy-2-[fluorine-18-fluoro-D-glucose ( $^{18}\text{F}$ -FDG). Most prevalently, maximum standardized uptake value (SUVmax) is reported as the staple marker of lesion metabolic activity level. Cut-off values based on this to differentiate malignant from benign lesions are often relayed in the literature. For instance  $\text{SUV} \geq 2.5$  is a widely adopted cut-off for malignancy [1, 2]. This remains common practice despite studies repeatedly questioning their general validity [3, 4] as SUV reliability and comparability is influenced by many factors [5]. Though SUVmax is simple to compute with high inter-observer reproducibility, it has shown to have a poorer inter-study repeatability than mean lesion uptake measures (SUVmean) [6-9]. As a single-voxel value, SUVmax is particularly susceptible to image noise and reconstruction parameters (e.g., smoothing, system response modelling, time-of-flight) with a shown noise level dependent bias [6, 10]. Variabilities in SUVmax of at least 10%-25% across different centres are reportedly expectable [7, 11]. How well SUVmax and cut-off values indicate malignancy is therefore questionable.

Alternative measures have been suggested, but are still only sporadically investigated. This includes the more robust SUVpeak [5, 10] as well as SUVmean and total lesion glycolysis (TLG) metrics that incorporate metabolic volume [12]. All depend on a region-of-interest (ROI) definition, which lacks standardized implementation, and are also more affected by partial-volume effect (PVE) [13]. Volumetric indices, nevertheless, offer accounts of overall  $^{18}\text{F}$ -FDG uptake avidity and extent, which might provide better disease characterization if measured correctly. This, however, is yet to be established and would have high clinical value in improving diagnostics, prognostics, and treatment planning.

**Table 1.** Variations in PET reconstruction parameters used in exams on the GE Discovery (D) systems.

System	GE D-STE	GED-VCT	GE D-RX	GE D-690
Matrix	6	9	9	1
Voxel size (mm <sup>3</sup> )	128×128	128×128	128×128	128×128
Algorithm	5.47×5.47×3.27	5.47×5.47×3.27	5.47×5.47×3.27	2.73×2.73×3.27
Iterations/subsets	3D OSEM	3D OSEM	3DOSEM	3DOSEM
Gaussian post-filter (mm FWHM)	2/28	2/28	2/21	3/18
Other	PSF modelling	PSF modelling	PSFmodelling	PSFmodelling+TOF

FWHM=full-widthathalf-maximum,PS=pointspreadfunction,TOF=time-of-flight

Measures to better grade disease severity and extent is furthermore especially desirable in response monitoring, where reliability of the standard single-voxel SUVmax can be compromised by its longitudinal instability. We therefore started systematic investigations of the utility of these semi-quantitative metrics in such applications, addressing here first the prevalent employment for static disease indication. This assessment of semi-quantitative classification considers the proposed use in management of incidental colorectal findings (ICF) as an exemplary high-impact clinical application. It is reported to occur in around 1%-4% of all patients subject to <sup>18</sup>F-FDG PET/CT [14-19] with a generally acknowledged clinical significance of focal ICF, e.g., novel malignancies [19-24]. Current recommended invasive follow-up assessment, such as colonoscopy [15, 25], could be avoided with PET lesion classification, which also entails the cross-study comparability challenges with scan variability typical of clinical settings. We evaluated such applied analysis using common methods available in commercial software.

## Subjects and Methods

Evaluation of quantitative assessment of soft tissue masses was conducted as a retrospective study on ICF in patients examined at our institution in 2011. Only those with colonoscopic results were included, which amounted to 25 patients (13 female, 12 male, mean age±SD: 71±12yr.) with a total of 30 focal lesions (8 cancers, 13 polyps/adenomas, 9 with only benign findings).

### <sup>18</sup>F-FDG PET/CT imaging

The scan exams all employed the institutional standard protocol based on the guidelines of the European Association of Nuclear Medicine [26]. Patients were asked to fast for at least six hours prior to the exam and given oral water as a negative contrast agent. Positron emission tomography/Computed tomography (PET/CT) was performed on four different integrated scanners (GE Discovery STE, VCT, RX,

and 690) and comprised: 1) Low-dose (<1mSv) CT scan (1-13mAs, 120 or 140kV, 30-87mA, 10 or 40mm total collimation width, pitch factor of 0.98-1.7, rotation time of 0.8s, table speed of 21.9-68.8mm/s, table feed of 17.5-55mm/rotation, 50cm field-of-view) for attenuation correction in all patients. The CT image matrix was 512×512 with a voxel size of 1.37×1.37×3.7mm<sup>3</sup>. 2) Standard dose diagnostic CT scan (same scanner parameters as for CT scan 1) but with a pitch factor of 0.98-1.85) acquired dependent on indication with-out or with intravenous contrast administration (Ultravist 370mg/mL iodine concentration, dose range: 40-80mL). The CT image matrix was 512×512 with a voxel size of 0.98×0.98×3.75mm<sup>3</sup>. 3) Subsequent PET scan acquired on average 73.8±14.3min after the injection of 4MBq/kg of <sup>18</sup>F-FDG (dose range: 198-402MBq). The scans covered the vertex or base of the skull down to the mid-thigh, with 6-8 bed positions of 2-3.5min each (depending on patient body mass index) and acquired in three-dimensional (3D) mode.

Positron emission tomography image reconstructions were with the GE Advance scanner software (GE Healthcare, Waukesha, WI, USA), employing a standard 3D iterative ordered subset expectation maximization (OSEM) algorithm with all corrections (attenuation, randoms, deadtime, and normalization) incorporated within the iterative loop. A model-based scatter correction was also applied to data and reconstruction parameters were as detailed in Table 1.

### Image analysis

To test ROI delineation variability and relate to literature reports, common approaches were applied with different segmentation techniques as implemented in a commercial dedicated software tool (ROVER, ABX, Radeberg, Germany). Here PET image intensities were converted to SUV normalized to body weight (BW) as the measure of distribution volume. For each dataset a volume encompassing the lesion and at least 2-3 voxels away from the lesion boundaries was manually defined. Subsequent analysis within this volume included automatic lesion detection with three segmentation methods for lesion SUV extraction and a built-in partial-volume correction (PVC):

- 1) Fixed 40% of SUVmax threshold (T40) based on guidelines [26] and reported use [16].
- 2) Adaptive background adjusted thresholds iteratively determined from an initial chosen isocontour at 40% of lesion SUVmax and calculated as [27]:

$$T = 0,39 \cdot (SUV_{max} - SUV_{bg}) + SUV_{bg} \quad (1)$$

where SUVbg is the mean uptake in a 3D shell of surrounding background with Eq. (1) evaluated globally for the entire lesion (LT method) or locally for each lesion relevant voxel (VT method).

- 3) Image-based PVC, where the lesion spill-out is estimated in a region,  $S$ , within a full-width at half-maximum (FWHM) distance from the estimated lesion (ROI) boundary. In this, the mean uptake, SUVspill, is found to yield the corrected lesion uptake as [28]:

$$cSUV_{mean} = SUV_{ROI} + SUV_{spill} = SUV_{ROI} + \frac{BW}{D} \cdot \sum_{v \in S} (A_s(v) - A_{bg}(v)) / V_{ROI} \quad (2)$$

where  $A_s$  is the measured activity and  $A_{bg}$  the corresponding background level of voxel  $v$  in the spill-out region,  $V_{ROI}$  the ROI volume,  $D$  the decay corrected tracer dose (MBq) and  $BW$  the body weight (g).

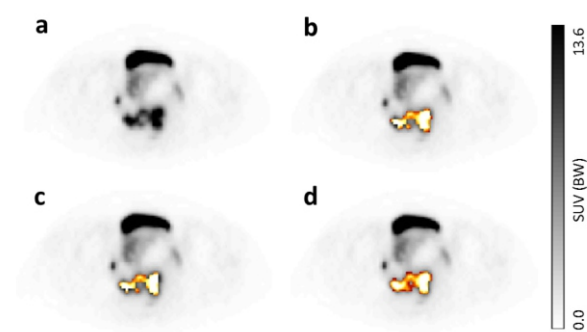
No lower limit cut-off value was used in segmentations to avoid potential inclusion of background related voxels, which were instead removed manually when apparent on visual inspection. Within the resulting ROI delineations measures of SUVmax, SUVmean, metabolically active volume (MAV), mean TLG = SUVmean MAV, and corresponding PVE corrected cSUVmean and cTLG were obtained. Additionally, SUVpeak values were also recorded following guideline recommendations [26] as the mean SUV in a fixed 1.2cm diameter spherical ROI centred at the voxel of SUVmax when such an area could be defined within lesion delineations.

### Statistical tests

Precursory comparisons between lesion groups (cancers, polyps/adenomas, benign findings) by Kruskal-Wallis test were performed and differences in derived semi-quantitative lesion measures were evaluated for statistical significance by nonparametric two-sided Wilcoxon rank sum test. Changes with PVC and method variability were assessed by Wilcoxon signed-rank matched-pairs test. All statistical analyses were carried out using MATLAB 9.0 (Mathworks Inc.) and a statistical significance level of 5%.

## Results

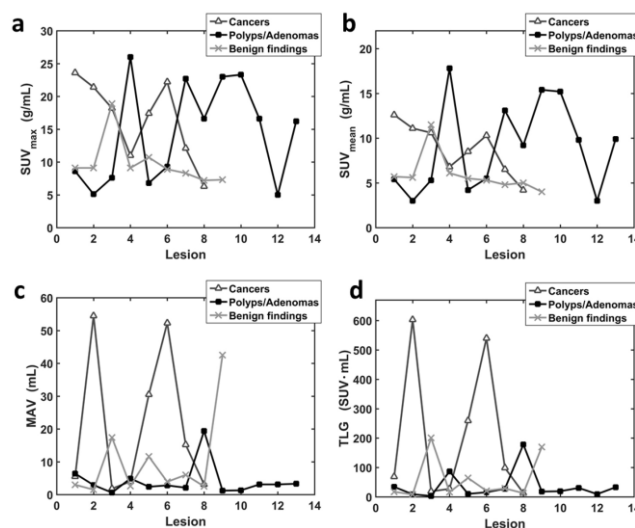
Positron emission tomography/CT imaging presented a varied appearance of focal ICF with no visually conspicuous characteristics discerning one histological lesion type from the other. Overall 30% (9/30) of these showed irregular shape and/or clear heterogeneity of uptake, which especially caused various degrees of variation in ROI definitions between segmentation methods as exemplified in Figure 1.



**Figure 1.** Axial image example of focal ICF uptake (a) and corresponding lesion delineations employing a fixed 40% of SUVmax threshold (b), adaptive lesion- (c) and voxel-based thresholds (d). Variations are seen between the different segmented ROI (overlaid in yellow-red colourmap).

### Quantitative measures

Across all lesions, SUVmax was  $>5$  and overlap in values was seen between histological groups (Figure 2a) with collectively indistinct levels ( $P=0.2$ ). Group results are given in Table 2. Though SUVmax in malignancies compared to benign values on average were significantly higher ( $P=0.03$ ) the levels of polyps/adenomas were indistinguishable from both of these groups ( $P=0.38-0.49$ ). Similarly, in all ROI delineations, SUVmean was  $\geq 2.7$  with no overall significant deviations between lesion groups ( $P=0.12-0.21$ ). With SUVmean on average higher in malignant than benign findings ( $P=0.01-0.04$ ), the measures in polyps/adenomas again obscured any distinction ( $P=0.27-0.77$ ) as is apparent in Figure 2b.



**Figure 2.** Semi-quantitative measures SUVmax (a), SUVmean (b), MAV (c), and TLG (d) extracted from lesion delineations using the VT segmentation method on each focal ICF in all three histological subgroups. Malignant and benign findings display overlap in SUV levels (a, b) and appear more distinct in volumetric measures (c, d). Only lesion 7 in the polyps group was imaged with time-of-flight.

Volumetric measures, on the other hand, displayed some clearer separation of malignancies from polyps/adenomas and benign findings (Figure 2c-d). This was predominantly due to noticeable MAV peaks in cancers well above the

**Table 2.** Results obtained with the three segmentation methods for all histological lesion groups.

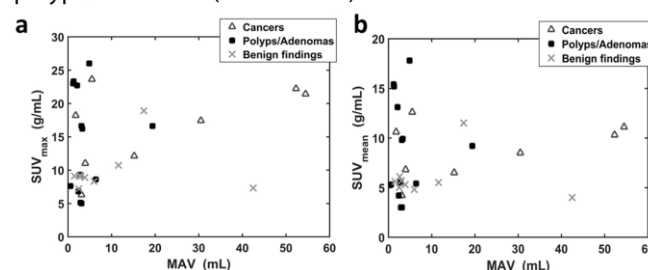
Measure	Method	Cancers (mean±SD)	Polyps/Adenomas (mean±SD)	Benign findings (mean±SD)
SUVmax	T40, LT, VT	16.5±6.2	14.4±7.7	9.8±3.6
SUVmean	T40	9.73±3.34	9.08±5.19	5.83±2.37
	LT	10.21±3.45	9.47±5.23	6.36±2.24
	VT	8.81±2.83	8.98±5.08	5.94±2.17
MAV (mL)	T40	14.74±14.82	3.98±3.92	10.27±12.31
	LT	12.76±13.26	3.41±3.68	6.36±6.03
	VT	20.86±22.18	4.12±4.84	10.11±13.22
TLG	T40	162.95±189.00	35.18±42.35	60.42±69.31
	LT	147.63±175.01	32.53±41.05	46.00±58.27
	VT	203.61±241.18	36.55±47.58	59.87±73.36

measures of the other two lesion types, which had similar ranges. However, fluctuating MAV levels made the groups on the whole statistically indiscernible ( $P=0.06$ ). While MAV varied with segmentation method (Table 2) the inter-group relations seen in Figure 2c were consistent across methods. Between groups, the larger malignant mean MAV were not significant compared to benign MAV ( $P=0.28-0.96$ ). Polyp/adenoma MAV, on the contrary, were generally smaller than both malignant (T40:  $P=0.06$ ; LT, VT:  $P=0.03-0.05$ ) and benign lesion MAVs (T40:  $P=0.05$ ; LT, VT:  $P=0.09-0.20$ ). Much of these volumetric relations carried over to TLG measures (Figure 2d, Table 2). Here, no significant relation was found in the average higher malignant values, above both benign ( $P=0.14-0.28$ ) and polyp/adenoma TLG ( $P=0.06$ ), which overall were lower than in benign findings ( $P=0.26-0.64$ ).

An assessment of the combined relation between SUV and estimated lesion MAV (Figure 3) revealed no obvious correlation between lesion size and extracted measures of  $^{18}\text{F}$ -FDG-uptake avidity. Although some lumping related to volumes of polyps/adenomas and benign findings appeared in the low range, half of the malignant lesions were seen to fall within these levels. With the spread in SUV, no further characteristic clustering among lesion types was clearly observed.

SUVpeak could not be defined for five lesions, mainly when MAV <2.5mL. For values recorded, similar group level relations as seen for SUVmax in Figure 2a were found for SUVpeak (not shown). Average peak levels were in malignant lesions  $12.9\pm4.5$  ( $n=7$ ), polyps/adenomas  $9.8\pm6.1$  ( $n=10$ ), and benign findings  $7.5\pm3.1$  ( $n=8$ ). Further paralleling SUVmax were overall indistinguishable group ranges ( $P=0.12$ ), where the distinction between malignant and benign

findings ( $P=0.03$ ) was obscured by the overlapping levels of polyps/adenomas ( $P=0.17-0.98$ ).



**Figure 3.** SUVmax (a) and SUVmean (b) as a function of the estimated metabolic active volume (MAV) of each focal ICF in all three histological subgroups measured with the VT segmentation method.

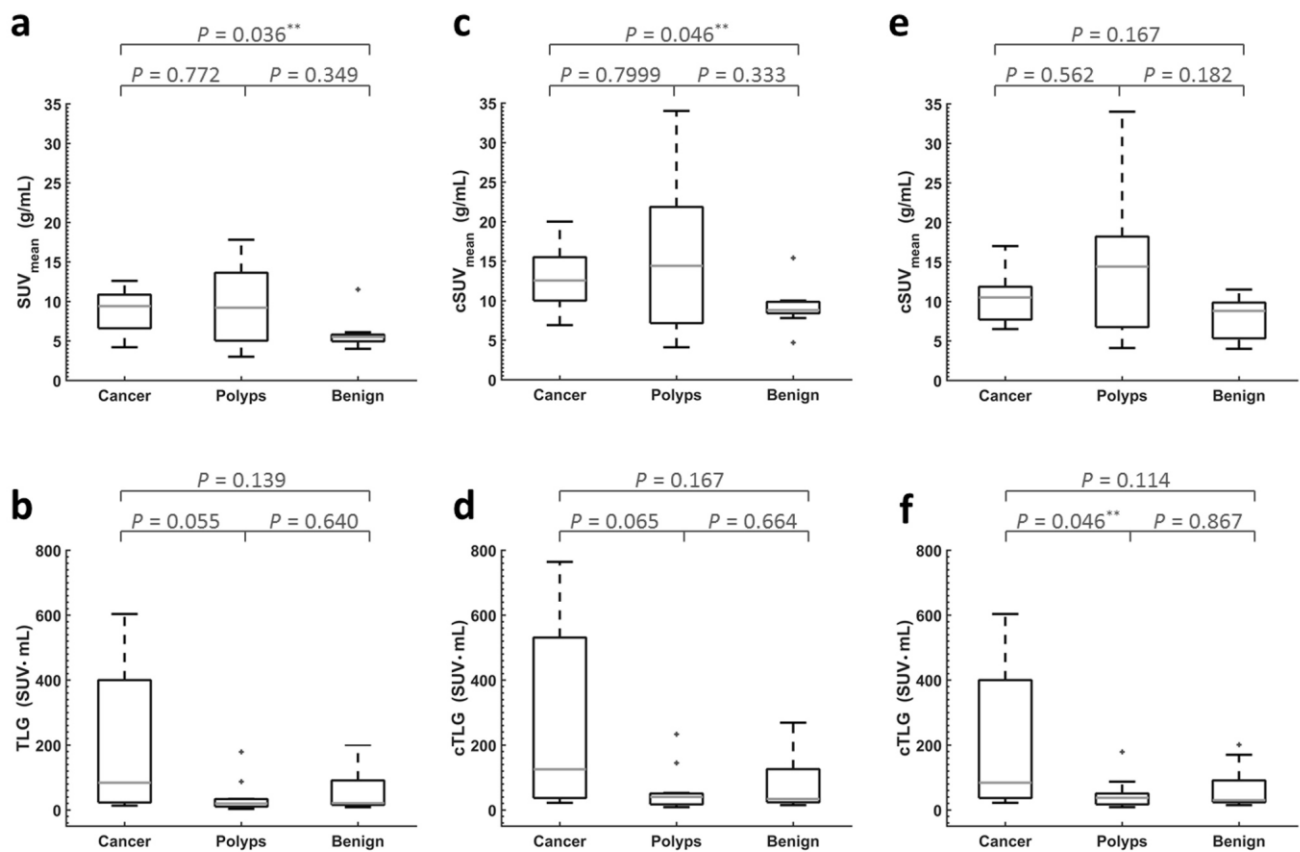
### Partial-volume correction

Application of the described PVC generally increased SUVmean and TLG measures significantly by a factor of 1.2-2.7 ( $P<0.008$ ). The corrections varied dependent on the segmentation technique. However, within each method, correction of values in all lesions amounted to mere shifts as relative group levels were largely preserved (Figure 4a-d). This remained the case when only correcting uptake measures in small lesions  $\leq 2\text{cm}$  (Figure 4e-f), where PVE typically is most pronounced. Hence no improvement in lesion type separability for SUVmean was achieved with the employed PVC, which also left the distinctions in TLG unchanged.

### Method variability

A 10%-70% of the threshold-based segmentations (T40: 3/30, LT: 8/30, VT: 21/30) gave improper lesion delineations





**Figure 4.** SUVmax (a) and SUVmean (b) as a function of the estimated metabolic active volume (MAV) of each focal ICF in all three histological subgroups measured with the VT segmentation method.

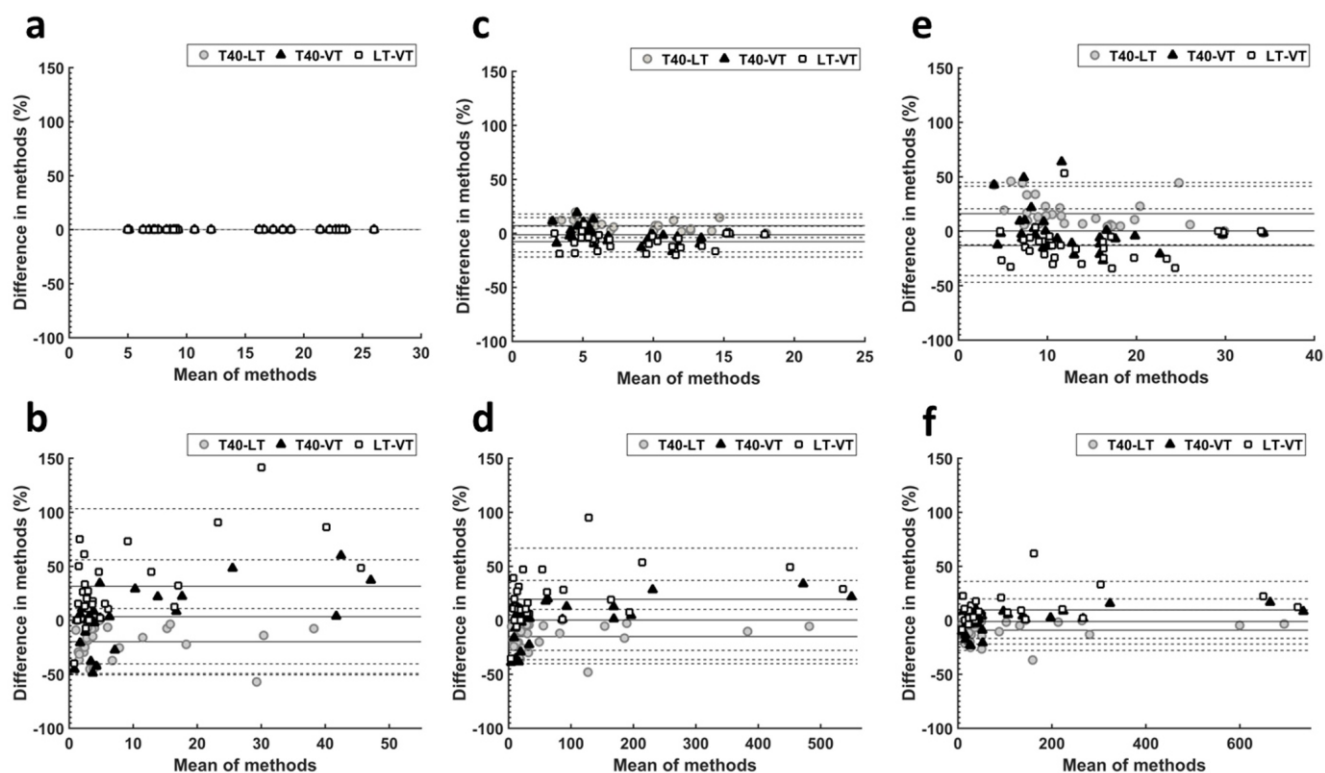
that included irrelevant voxels, which had to be manually excluded. Across all these segmentations, SUVmax was reproducible (Figure 5a). On the other hand, MAV measurement varied the most (Figure 5b) between -58.6%-141.5% among delineation techniques. This translated into inter-method variations of -20.2%-25.2% in SUVmean and -48.7%-94.9% in TLG (Figure 5c-d). Irrespective of these differences in measured values the lesion group level relations shown in Figure 2 were overall consistently found across methods. The further impact of PVC for each segmentation did not yield a larger between-method agreement in measurements. In fact, increased spread in method differences of -38.9%-63.6% in cSUVmean was recorded (Figure 5e), albeit with collectively decreased variations of -38.2%-61.8% in cTLG (Figure 5f). In none of the ROI delineation approaches did these corrected uptake estimates overall alter the corresponding uncorrected relations seen in Figure 2. On the whole, LT method measures, without and with PVC, deviated significantly ( $P < 0.02$ ) from the more concordant results of the T40 and VT methods.

## Discussion

Within common clinical settings and procedures, analysis of PET scans of ICF clearly showed SUVmax to be an insuff-

icient marker of aetiology (i.e., malignant, polyp/adenoma, or benign). The displayed overlap between lesion types echoes multiple reports in the literature [16, 21, 23, 29-32], where uptake levels were also noted to be inconsistent between studies (Figure 6). Although some report positive discriminatory qualities [19, 32-34], a seemingly larger number writes off SUVmax as a stand-alone tool for differentiating histology [14, 17, 18, 25, 29]. The further collective lack of proven reproducibility of favourable findings and high inter-study variability seen in the literature speaks to an absence of general validity. Given the broad test-retest vulnerability seen in SUVmax even with harmonized procedures [9], this altogether substantiates that SUVmax cannot casually be relied on for accurate indication of malignancy.

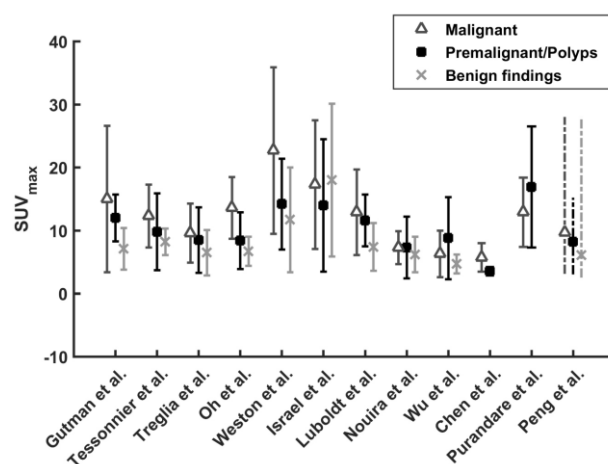
Reported reference values are then also more likely study-dependent, particularly cut-off suggested to signify malignancy. As is apparent from the present study, any SUVmax cut-off defined will not be clinically significant. The widely reported overlap in uptake levels verified here would inevitably entail mislabelling with cut-offs based on SUVmax alone. Such inapplicability and the discordant reports are seen among studies of both small (<50) [19, 23, 25, 29] and larger sample sizes (>100) [14, 18, 34]. It is therefore in all probability measurement variability rather than histopathological significance that marks SUVmax cut-off definitions as also evident from inconsistent values between 2.5-11.4 proposed in the literature [14, 16, 18, 31-33]. In fact, this ap-



**Figure 5.** Bland-Altman plots of relative differences in measured SUVmax (a), MAV (b), SUVmean (c), TLG (d), and PVE corrected cSUVmean (e) and cTLG (f) between the three segmentation methods (T40, LT, VT), respectively. Solid and dashed lines indicate mean  $\pm$  2SD.

pears to apply generally to reports of such cut-off for malignancy with a divergence in values within and across soft tissue pathologies (Table 3). Certainly, the notion of a universally significant cut-off, such as 2.5, is arbitrary and must be dismissed. Diversities in technical aspects, procedures and methodology typical of clinical settings among facilities imply that suggested cut-off values prevalently are at best centre-dependent. Results on ICF in this study and others represent well the practice and outcome variability one might expect in these scenarios, showing limitations to the use of SUVmax and derived cut-off values. Such applications may improve with method standardisations as broadly called for. However, the degree of test-retest repeatability remains an issue that must be accounted for in clinical trials.

SUVpeak suggested for increased inter-study robustness was no more discriminatory than SUVmax in this study. Its application furthermore was limited by and would depend on the ROI definition, which can vary in implementation. Hence SUVpeak will not be more reliable for aetiological differentiation than SUVmax or SUVmean as it ranges somewhere between the two. In small lesions, SUVpeak tends towards being a ROI-specific SUVmean measure, which across delineation methods did not provide adequate distinctions either in this study. Relative SUVmean levels between lesion groups actually resembled those seen in SUVmax. Such could allude to a correspondence between the two as has also been found previously [32], perhaps merely reflecting the general grade of SUV measurements and predominant levels in ICF uptake distributions.



**Figure 6.** Literature reports on mean SUVmax and standard deviations (bars) or ranges (dashed lines) for different categories of focal ICFs in  $^{18}\text{F}$ -FDG PET/CT.

Evaluation of alternative volumetric measures, MAV and TLG, showed somewhat clearer separation of malignant and non-malignant lesions. This is contrary to another study reporting less discriminatory efficacy of MAV than SUVmax [32], which may be due to study dissimilarities. That MAV in this study nonetheless yielded better distinction of particularly polyps/adenomas whereas SUVs overall were only distinct between malignant and benign findings resembles that reported by Oh et al. (2012) [16]. Such would

**Table 3.** Reported SUVmax cut-off values for malignancy in various soft tissue masses

Type	First author (year)	Threshold definition	Threshold (s)
Colo-rectal findings	Weston (2010) [31]	True possitive cut-off	3.5
	Luboldt (2010) [32]	Best cut-off	5.0
	Peng (2011) [14]	Uptake-positive cut-off	2.5
	Oh (2012) [16]	Best cut-off	9.1
	Cho (2013) [33]	Bestcut-off	5.05
	van Hoeij (2015) [18]	Best cut-off	11.4
Non-small cell lung cancer	Jeong (2002) [36]	Best cut-off	7.0
	Sasaki (2003) [37]	Best cut-off	5.0
	Van Baardwijk (2007) [38]	Best cut-off	8.0/11.0
	Al-Sarraf (2008) [39]	Best cut-off	15.0
Thoracic lymph node metastases	Bryant (2006) [40]	Best cut-off	5.3
	Tournoy (2007) [41]	Best cut-off	2.9
	Bella (2014) [42]	Best cut-off	4.1
	Bryant (2006) [40]	Diagnostic criteria	2.5
	Tasci (2010) [43]	Diagnostic criteria	5.2
Pulmonary nodes	Degirmenci (2008) [44]	Best cut-off	2.4
	Grgic (2010) [45]	Best cut-off	4.0
	Huang (2012) [46]	Best cut-off	5.05
	Lopez (2015) [47]	Best cut-off	1.95

Adrenal lesions	Okada (2009) [2]	Best cut-off	2.5
	Ozcan (2011) [48]	Best cut-off	4.2
	Kunikowska (2014) [49]	Best cut-off	5.2
	Launay (2015) [50]	Best cut-off	3.7

indicate a better differentiation based on combined measures rather than on any single index. Although their proposed compounded two-fold cut-off approach seems promising some misclassification occurred. Our study found no obvious correlation between MAV and SUV. Instead, the overlaps between subgroups found here (Figure 3) would imply a risk of false labelling for any defined set of cut-off. It is still further clear that such combined cut-offs will be just as study-dependent as a unilateral cut-off might be given the variation in measurements observed. Even so, integrating both activity extent and level appears to offer greater characteristics since the blended TLG measure alone showed better distinction of malignancy than either SUV or MAV in our study. TLG in this context has apparently been little investigated to date as we found only one recent study reporting on TLG for ICF classification [35]. A favourable application of MAV and TLG was recounted there too.

Irrespective of the applied semi-quantitative index none of the ICF studies in the literature we reviewed dealt explicitly with the potential biases induced by PVE. Considering the range of reported ICF sizes many cases would expectedly be affected. As PVE is study and anatomy dependent such unaccounted errors stand for additional variability particularly in SUVmean and TLG estimates. This can explain some fluctuations in these measures for lesion subgroups. Possible corrections with an image-based technique yielded shifted levels but overall unaltered relations. A confirmation of observed relative levels was, however, not necessarily obtained by that. It perhaps more likely reflects that the image-based PVC method employed is too simple and inadequate as corrections were also shown to be segmentation-dependent. As PVE is the same in any given lesion, its correct estimation and correction should be invariable with the applied methods. Nonetheless, the significant correction factors of 1.2-2.7 computed in this study can indicate that PVE in lesions like these is non-negligible as expected. Hence results in this regard remain inconclusive. More proper PVC techniques are called for to reliably assess the utility of the affected SUV-based measures. PVE must not at any rate be left out of the equation, as has been the case so far.

The threshold-based segmentation methods proved insufficient with variable results, here often requiring manual processing that increase variability. It can further be inferred from Figure 5 that different ROI definitions also has a direct impact on absolute quantitative results. Therefore, any



cut-off value based on ROI-dependent measures will clearly not have general validity if different segmentation approaches are employed. The variability is clearly seen in the highly method-dependent volume estimates with the largest variation among all measures of up to 141.5%. However, the fact that the corresponding SUVmean and TLG measures differ less among methods signifies the inherent robustness in applying region averaging. Such would also underlie reported higher inter-study reproducibility compared to SUVmax. Of course SUVmean variability would increase with the degree of heterogeneity, where alternatively SUVmax is rendered even more unlikely to reflect overall metabolism well. In this study, different segmentations yielded different biases in absolute measurements, but similar relative levels between lesion subgroups were found irrespective of method. This might just reflect that the within-method bias is the same and inhomogeneity was only to such a degree that differences in biases between methods then amount to relative value shifts. More accurate PVC would expectedly increase the consistency in TLG and SUVmean measurements across delineation methods.

Inter-study standardization will heighten comparability of uptake measures, but is limited still by practical variability and the degree of repeatability. In this study, the inter-method shifts in scale with overall consistent lesion uptake relations further imply that relative rather than absolute scale-dependent comparisons should be preferred. In any respect, proper ROI definition matters, especially with highly inhomogeneous uptake. Thus, improved analyses with more robust and reliable segmentations and PVC need to be developed to accurately assess the utility of extended ROI measures for better disease characterization.

In conclusion, this study of ICF demonstrated that the prevalently used SUVmax measure is insufficient for aetiological differentiation of lesions. No useful concomitant cut-off value for malignancy could be defined as such often reported values also will be highly study-dependent. Semi-quantitative analysis with common methods showed that alternative measures such as SUVpeak and SUVmean are no better as single markers. A combination of measures rather than any particular index might provide better classification as the compounded TLG alone indicated larger histological distinctions by integrating metabolic volume and activity level. The variability of absolute uptake estimates and consistent relative levels seen with different segmentations and PVC depend on the accuracy of both methods. As the common threshold-based segmentation and simple PVC methods employed in this study proved sub-optimal it remains to be assessed whether improved methods can provide measures, e.g., cTLG, that better reflect the disease severity and extent than SUVmax values.

The authors of this study declare no conflict of interest

## Bibliography

- Lucignani G, Paganelli G, Bombardieri E. The use of standardized uptake values for assessing FDG uptake with PET in oncology: a clinical perspective. *Nucl Med Commun* 2004; 25(7): 651-6.
- Okada M, Shimono T, Komeya Y et al. Adrenal masses: the value of additional fluorodeoxyglucose-positron emission tomography/computed tomography (FDG-PET/CT) in differentiating between benign and malignant lesions. *Ann Nucl Med* 2009; 23(4): 349-54.
- Basu S, Zaidi H, Salavati A et al. FDG PET/CT methodology for evaluation of treatment response in lymphoma: from "graded visual analysis" and "semiquantitative SUVmax" to global disease burden assessment. *Eur J Nucl Med Mol Imaging* 2014; 41(11): 2158-60.
- Houshmand S, Salavati A, Hess S et al. An update on novel quantitative techniques in the context of evolving whole-body PET imaging. *PET clinics* 2015; 10(1): 45-58.
- Adams MC, Turkington TG, Wilson JM, Wong TZ. A systematic review of the factors affecting accuracy of SUV measurements. *Am J Roentgenol* 2010; 195(2): 310-20.
- Krak NC, Boellaard R, Hoekstra OS et al. Effects of ROI definition and reconstruction method on quantitative outcome and applicability in a response monitoring trial. *Eur J Nucl Med Mol Imaging* 2005; 32(3): 294-301.
- Fahey FH, Kinahan PE, Doot RK et al. Variability in PET quantitation within a multicenter consortium. *Med Phys* 2010; 37(7): 3660-6.
- Nahmias C, Wahl LM. Reproducibility of standardized uptake value measurements determined by  $^{18}\text{F}$ -FDG PET in malignant tumors. *J Nucl Med* 2008; 49(11): 1804-8.
- de Langen AJ, Vincent A, Velasquez LM et al. Repeatability of  $^{18}\text{F}$ -FDG uptake measurements in tumors: a metaanalysis. *J Nucl Med* 2012; 53(5): 701-8.
- Boellaard R, Krak NC, Hoekstra OS, Lammertsma AA. Effects of noise, image resolution, and ROI definition on the accuracy of standard uptake values: a simulation study. *J Nucl Med* 2004; 45(9): 1519-27.
- Iskender I, Kadioglu SZ, Kosar A et al. Is there any maximum standardized uptake value variation among positron emission tomography scanners for mediastinal staging in non-small cell lung cancer? *Interact Cardiovasc Thorac Surg* 2011; 12(6): 965-9.
- Larson SM, Erdi Y, Akhurst T et al. Tumor Treatment Response Based on Visual and Quantitative Changes in Global Tumor Glycolysis Using PET-FDG Imaging. The Visual Response Score and the Change in Total Lesion Glycolysis. *Clin Positron Imaging* 1999; 2(3): 159-71.
- Soret M, Bacharach SL, Buvat I. Partial-volume effect in PET tumor imaging. *J Nucl Med* 2007; 48(6): 932-45.
- Peng J, He Y, Xu J et al. Detection of incidental colorectal tumours with  $^{18}\text{F}$ -labelled 2-fluoro-2-deoxyglucose positron emission tomography/computed tomography scans: results of a prospective study. *Colorectal Dis* 2011; 13(11): e374-e8.
- Lin M, Koo JH, Abi-Hanna D. Management of Patients Following Detection of Unsuspected Colon Lesions by PET Imaging. *Clin Gastroenterol Hepatol* 2011; 9(12): 1025-32.
- Oh J-R, Min J-J, Song H-C et al. A stepwise approach using metabolic volume and SUVmax to differentiate malignancy and dysplasia from benign colonic uptakes on  $^{18}\text{F}$ -FDG PET/CT. *Clin Nucl Med* 2012; 37(6): e134-e40.
- Treglia G, Taralli S, Salsano M et al. Prevalence and malignancy risk of focal colorectal incidental uptake detected by  $^{18}\text{F}$ -FDG-PET or PET/CT: a meta-analysis. *Radiol Oncol* 2014; 48(2): 99-104.
- van Hoeij FB, Keijsers RG, Loffeld BC et al. Incidental colonic focal FDG uptake on PET/CT: can the maximum standardized uptake value (SUVmax) guide us in the timing of colonoscopy? *Eur J Nucl Med Mol Imaging* 2015; 42(1): 66-71.
- Fuertes J, Montagut C, Bullich S et al. Incidental focal uptake in colorectal location on oncologic  $^{18}\text{F}$ FDG PET and PET/CT studies: histopathological findings and clinical significances. *Rev Esp Med Nucl Imagen Mol* 2015; 34(2): 95-101.
- Kamel EM, Thumshirn M, Truninger K et al. Significance of incidental  $^{18}\text{F}$ -FDG accumulations in the gastrointestinal tract in PET/CT: correlation with endoscopic and histopathologic results. *J Nucl Med* 2004; 45(11): 1804-10.
- Israel O, Yefremov N, Bar-Shalom R et al. PET/CT detection of unexpected gastrointestinal foci of  $^{18}\text{F}$ -FDG uptake: incidence, localization patterns, and clinical significance. *J Nucl Med* 2005; 46(5): 758-62.
- Lee JC, Hartnett GF, Hughes BG, Ravi Kumar AS. The segmental distribution and clinical significance of colorectal fluorodeoxyglucose uptake incidentally detected on PET-CT. *Nucl Med Commun* 2009; 30(5): 333-7.
- Treglia G, Calcagni ML, Rufini V et al. Clinical significance of incidental focal colorectal  $^{18}\text{F}$ -fluorodeoxyglucose uptake: our experience and a

- review of the literature. *Colorectal Dis* 2012; 14(2): 174-80.
24. Zhuang H, Hickeson M, Chacko TK et al. Incidental detection of colon cancer by FDG positron emission tomography in patients examined for pulmonary nodules. *Clin Nucl Med* 2002; 27(9): 628-32.
  25. Liu T, Behr S, Khan S et al. Focal Colonic FDG Activity with PET/CT: Guidelines for Recommendation of Colonoscopy. *World J Nucl Med* 2015; 14(1): 25-30.
  26. Boellaard R, O'Doherty MJ, Weber WA et al. FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. *Eur J Nucl Med Mol Imaging* 2010; 37(1): 181-200.
  27. Hofheinz F, Langner J, Petr J et al. An automatic method for accurate volume delineation of heterogeneous tumors in PET. *Med Phys* 2013; 40(8): 082503.
  28. Hofheinz F, Langner J, Petr J et al. A method for model-free partial volume correction in oncological PET. *Eur J Nucl Med Mol Imaging Res* 2012; 2(1): 16.
  29. Gutman F, Alberini JL, Wartski M et al. Incidental colonic focal lesions detected by FDG PET/CT. *Am J Roentgenol* 2005; 185(2): 495-500.
  30. Tessonier L, Gonfrier S, Carrier P et al. Unexpected focal bowel 18-FDG uptake sites: should they be systematically investigated? *Bull Cancer* 2008; 95(11): 1083-7.
  31. Weston BR, Iyer RB, Qiao W et al. Ability of integrated positron emission and computed tomography to detect significant colonic pathology. *Cancer* 2010; 116(6): 1454-61.
  32. Luboldt W, Volker T, Wiedemann B et al. Detection of relevant colonic neoplasms with PET/CT: promising accuracy with minimal CT dose and a standardised PET cut-off. *Eur Radiol* 2010; 20(9): 2274-85.
  33. Cho SH, Kim SW, Kim WC et al. Incidental focal colorectal <sup>18</sup>F-fluorodeoxyglucose uptake on positron emission tomography/computed tomography. *World J Gastroenterol* 2013; 19(22): 3453.
  34. Shim JH, O JH, Oh SI et al. Clinical significance of incidental colonic <sup>18</sup>F-FDG uptake on PET/CT images in patients with gastric adenocarcinoma. *J Gastrointest Surg* 2012; 16(10): 1847-53.
  35. Suzuki Y, Okabayashi K, Hasegawa H et al. The association between metabolic parameters in FDG-PET/CT and pathological findings of colorectal cancer. *Gut* 2015; 64(Suppl 1): A564.
  36. Jeong H-J, Min J-J, Park J et al. Determination of the prognostic value of [<sup>18</sup>F] fluorodeoxyglucose uptake by using positron emission tomography in patients with non-small cell lung cancer. *Nucl Med Commun* 2002; 23(9): 865-70.
  37. Sasaki R, Komaki R, Macapinlac H et al. [<sup>18</sup>F] fluorodeoxyglucose uptake by positron emission tomography predicts outcome of non-small-cell lung cancer. *J Clin Oncol* 2005; 23(6): 1136-43.
  38. van Baardwijk A, Dooms C, van Suylen RJ et al. The maximum uptake of <sup>18</sup>F-deoxyglucose on positron emission tomography scan correlates with survival, hypoxia inducible factor-1α and GLUT-1 in non-small cell lung cancer. *Eur J Cancer* 2007; 43(9): 1392-8.
  39. Al-Sarraf N, Gately K, Lucey J et al. Clinical implication and prognostic significance of standardised uptake value of primary non-small cell lung cancer on positron emission tomography: analysis of 176 cases. *Eur J Cardiothorac Surg* 2008; 34(4): 892-7.
  40. Bryant AS, Cerfolio RJ, Klemm KM, Ojha B. Maximum standard uptake value of mediastinal lymph nodes on integrated FDG-PET-CT predicts pathology in patients with non-small cell lung cancer. *Ann Thorac Surg* 2006; 82(2): 417-23.
  41. Tournoy K, Maddens S, Gosselin R et al. Integrated FDG-PET/CT does not make invasive staging of the intrathoracic lymph nodes in non-small cell lung cancer redundant: a prospective study. *Thorax* 2007; 62(8): 696-701.
  42. Bella AJE, Zhang Y-R, Fan W et al. Maximum standardized uptake value on PET/CT in preoperative assessment of lymph node metastasis from thoracic esophageal squamous cell carcinoma. *Chin J Cancer* 2014; 33(4): 211-7.
  43. Tasci E, Tezel C, Orki A et al. The role of integrated positron emission tomography and computed tomography in the assessment of nodal spread in cases with non-small cell lung cancer. *Interact Cardiovasc Thorac Surg* 2010; 10(2): 200-3.
  44. Degirmenci B, Wilson D, Laymon CM et al. Standardized uptake value-based evaluations of solitary pulmonary nodules using F-18 fluorodeoxyglucose-PET/computed tomography. *Nucl Med Commun* 2008; 29(7): 614-22.
  45. Grgic A, Yüksel Y, Gröschel A et al. Risk stratification of solitary pulmonary nodules by means of PET using <sup>18</sup>F-fluorodeoxyglucose and SUV quantification. *Eur J Nucl Med Mol Imaging* 2010; 37(6): 1087-94.
  46. Huang Y-E, Lu H-I, Liu F-Y et al. Solitary pulmonary nodules differentiated by dynamic F-18 FDG PET in a region with high prevalence of granulomatous disease. *J Radiat Res* 2012; 53(2): 306-12.
  47. van Gómez López O, García Vicente A, Honguero Martínez AF et al. <sup>18</sup>F-FDG-PET/CT in the assessment of pulmonary solitary nodules: comparison of different analysis methods and risk variables in the prediction of malignancy. *Transl Lung Cancer Res* 2015; 4(3): 228-35.
  48. Ozcan Kara P, Kara T, Kara Gedik G et al. The role of fluorodeoxyglucose-positron emission tomography/computed tomography in differentiating between benign and malignant adrenal lesions. *Nucl Med Commun* 2011; 32(2): 106-12.
  49. Kunikowska J, Matyskiel R, Toutouchi S et al. What parameters from <sup>18</sup>F-FDG PET/CT are useful in evaluation of adrenal lesions? *Eur J Nucl Med Mol Imaging* 2014; 41(12): 2273-80.
  50. Launay N, Silvera S, Tenenbaum F et al. Value of <sup>18</sup>F-FDG PET/CT and CT in the Diagnosis of Indeterminate Adrenal Masses. *Int J Endocrinol* 2015; 2015: 8.